PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE HONORABLE BOARD OF PATENT APPEALS AND INTERFERENCES

In re the Application of

Francine R. Chen et al.

Application No.: 10/626,875 Examiner: Burke, Jeff A.

Filed: July 25, 2003 Docket No.: A3053-US-NP
XERZ 2 01564

Title: SYSTEMS AND METHODS FOR LINKED EVENT DETECTION

BRIEF ON APPEAL

Appeal from Group 2165

Mark S. Svat, Esq., Reg. No. 34,261
FAY SHARPE LLP
The Halle Building – Fifth Floor
1228 Euclid Avenue
Cleveland, Ohio 44115
Telephone: (216) 363-9000
Attorneys for Appellants

CERTIFICATE OF TRANSMISSION

I hereby certify that this correspondence (and any item referred to herein as being attached or enclosed) is (are) being transmitted to the USPTO by electronic transmission via the EFS Web on the date indicated below.

9-10-09 Karen Forsyth

Date Name: Karen M. Forsyth

N:\XERZ\201564\ALB0000014V001.docx

TABLE OF CONTENTS

I. REAL PARTY IN INTEREST

The real party in interest for this appeal and the present application is Palo Alto Research Center Inc. (3333 Coyote Hill Rd., Palo Alto, California 94304), by way of an Assignment recorded in the U.S. Patent and Trademark Office at Reel 014333, Frame 0512.

II. RELATED APPEALS AND INTERFERENCES

There are no prior or pending appeals, interferences or judicial proceedings, known to Appellant, Appellant's representative, or the Assignee, that may be related to, or which will directly affect or be directly affected by or have a bearing upon the Board's decision in the pending appeal.

There is, however, a provisional rejection of claims 1, 20, 39, 58, and 81-84 on the ground of non-statutory obviousness-type double patenting over claims 1, 16, and 31-32 of co-pending Application No. 10/626,856 which has since been granted as U.S. Patent No. 7,577,654 and contains 14 claims total. Applicants consider the provisional rejection inoperable due to the issuance of the patent and the change in scope between the claims in the published application and the claims in the issued patent.

STATUS OF CLAIMS

Claims 1-88 are on appeal.

Claims 1-88 are pending.

Claims 1-88 are rejected.

III. STATUS OF AMENDMENTS

An Amendment After Final Rejection was filed on January 21, 2008. By an Advisory Action dated February 19, 2008, it was indicated that the requested amendments were not entered. Therefore, the claims presented for appeal are those as set forth prior to Applicants' Amendment After Final submitted January 21, 2008.

IV. SUMMARY OF CLAIMED SUBJECT MATTER

The invention of claim 1 is directed to a computer-implemented method of determining predictive models for a linked event detection system as shown in FIG. 2, including: determining source-identified training stories, determining inter-story similarity vectors in a memory **20** (FIG. 4) for at least one story-pair of the source-identified training stories, determining link label information for the at least one story-pair, and determining and storing at least one predictive model in the memory based on the inter-story similarity vectors and the link label information. The link label information indicates the existence of at least one link between a pair of stories in the source-identified training stories and that the linked source-identified stories are related to the same event.

A processor **15**, as shown in FIG. 4 of the specification, performs the step of determining source-identified training stories and the step of determining link label information as described on page 20, line 20 – page 21, line 3. The step for determining inter-story similarity vectors is performed by an inter-story similarity determining circuit **40** as described on page 21, line 28 – page 22, line 30. The step for determining and storing at least one predictive model is performed by a predictive model determining circuit **50** as described on page 20, lines 20-33.

The invention of claim 2 is directed to the computer-implemented method of claim 1, further including, as shown in FIG. 2, determining an inter-story similarity metric for the story-pairs, and determining source-pair statistics for the story-pairs for the inter-story similarity vectors recited in claim 1.

The invention of claim 20 is directed to a linked event detection training system that includes an input/output circuit **10** (FIG. 4), a memory **20**, and a processor **15**, as shown in FIG. 4, that receives source-identified training stories and associated link label information for at least one story-pair via the input/output circuit. This recited limitation is described starting on page 20, line 31, of the original specification, and continuing to page 31, line 3, where a user of personal computer **300** as shown in FIG. 4 initiates a request over communications link **99** to the link detection system **100** to determine a predictive model for the link detection system **100**. The request is received by the link detection system **100**, and the processor **15** activates the input/output circuit 10 to retrieve the source-identified stories **1000-1002** over communications link 99 and store the source-identified stories in memory **20**. As also recited in the claim, and described on page 5, lines 17-19, the link label information indicates the existence of at least one link between a pair of stories in the source-identified training stories and that the linked source-identified stories are related to the same event. Also included in the system is an inter-story similarity vector determining circuit **40, 45** that determines inter-story similarity vectors in the memory for at least one story-pair of the source-identified training stories as described from page 3, line 32, through page 4, line 4, inter-story similarity metrics are determined and source-pair similarity statistics are also determined and used to normalize the inter-story similarity metrics. The inter-story similarity metrics and the normalized source-pair similarity statistics are combined to form a similarity vector for each pair of stories. As further described on page 21, line 28, through page 22, line 30, the similarity metric component of the similarity vectors is determined by circuit **40**. Exemplary similarity metrics are described such as the Hellinger metric, the Tanimoto metric, the clarity-distance metric, and the cosine-distance metric. The source pair statistics component of

the similarity vectors is determined by the similarity statistics determining circuit **45**. Exemplary statistics described include the running median, mean, variance and the like. Further included is a predictive model determining circuit **50** that determines and stores at least one predictive model in the memory based on the inter-story similarity vectors and the link label information. This recited feature of determining a predictive model is discussed in further detail on page 22, lines 4-11, where exemplary use of a support vector machine, a decision tree inducer, a classifier or any other known method is described.

The invention of claim 21 is directed to the linked event detection training system of claim 20, further including, as shown in FIG. 2, a similarity metric determining circuit that determines an inter-story similarity metric for the story-pairs, and a similarity statistics determining circuit that determines source-pair statistics for the story-pairs for the inter-story similarity vectors recited in claim 20.

The invention of claim 39 is directed to a computer-implemented method of linked event detection as shown in FIG. 3, including: determining source-identified training stories, determining inter-story similarity vectors in a memory **20** (FIG. 4) for at least one story-pair of the source-identified training stories, determining at least one predictive model in the memory for link detection, determining a link between the story-pairs based on the predictive model and the inter-story similarity vector, and displaying the link on a computer **300** or storing the link in an information repository **200**. The link indicates that the story-pairs are related to the same event.

A processor **15**, as shown in FIG. 4 of the specification, performs the step of determining source-identified training stories as described on page 20, line 20 – page 21, line 3. The step for determining inter-story similarity vectors is performed by an inter-story similarity determining circuit **40** as described on page 21, line 28 – page 22, line

30. The step for determining at least one predictive model is performed by a predictive model determining circuit **50** as described on page 20, lines 20-33. The step for determining a link between the story-pairs is performed by a link determining circuit **55** as described on page 22, lines 20-30. The step for displaying the link on a computer or storing the link in an information repository is performed by the link determining circuit.

The invention of claim 40 is directed to the computer-implemented method of claim 39, further including, as shown in FIG. 3, determining an inter-story similarity metric for each story-pair, and determining source-pair statistics for the story-pairs for the inter-story similarity vectors recited in claim 39.

The invention of claim 58 is directed to a linked event detection system that includes an input/output circuit **10** (FIG. 4), a memory **20**, and a processor **15**, as shown in FIG. 4, that receives source-identified stories via the input/output circuit. This recited limitation is described starting on page 20, line 31, of the original specification, and continuing to page 31, line 3, where a user of personal computer **300** as shown in FIG. 4 initiates a request over communications link **99** to the link detection system **100** to determine a predictive model for the link detection system **100**. The request is received by the link detection system **100**, and the processor **15** activates the input/output circuit 10 to retrieve the source-identified stories **1000-1002** over communications link 99 and store the source-identified stories in memory **20**. An inter-story similarity vector determining circuit **40, 45** determines inter-story similarity vectors in the memory for the story-pairs of the source-identified stories as described from page 3, line 32, through page 4, line 4, inter-story similarity metrics are determined and source-pair similarity statistics are also determined and used to normalize the inter-story similarity metrics. The inter-story similarity metrics and the normalized source-pair similarity statistics are combined to form

a similarity vector for each pair of stories. As further described on page 21, line 28, through page 22, line 30, the similarity metric component of the similarity vectors is determined by circuit **40**. Exemplary similarity metrics are described such as the Hellinger metric, the Tanimoto metric, the clarity-distance metric, and the cosine-distance metric. The source pair statistics component of the similarity vectors is determined by the similarity statistics determining circuit **45**. Exemplary statistics described include the running median, mean, variance and the like. A link determining circuit **55** determines and displays on a computer **300** or stores in an information repository **200**, links between story-pairs based on a predictive model in the memory and the inter-story similarity vectors. The links indicate that the story-pairs are related to the same event. Operation of the link determining circuit **55** is described in more detail on page 22, lines 21-30. The recited predictive model is discussed in further detail on page 22, lines 4-11, where the predictive model is determined based on link label information, the similarity metrics and the source-pair similarity statistics for each story-pair.

The invention of <u>claim 59</u> is directed to the linked event detection system of claim 58, further including, as shown in FIG. 3, a similarity metric determining circuit that determines an inter-story similarity metric for the story-pairs, and a similarity statistics determining circuit that determines source-pair statistics for the story-pairs for the inter-story similarity vectors recited in claim 58.

The invention of <u>claim 77</u> is directed to a method of determining a stopword list as shown in FIG. 5. The method includes determining a source-identified training corpus of text information as described on page 23, lines 7-10, where the training corpus may be created from automatically recognized speech utterances transcribed to text, text files from a digital library, HTML or web pages served by a web server, or any known or later

developed information source. The method further includes determining a verified first source-mode transformation of the source-identified training corpus text from a first mode to a second mode based on a verified transcription or a verified translation. This recited feature is described on page 23, lines 11-18, with various exemplary transformations described. The claim also recites determining an un-verified second source-mode transformation of the source-identified training corpus text from a first mode to a second mode as described on page 23, lines 20-25, and determining at least one transformation error associated with distribution differences between the first and second transformations and identified sources as described on page 23, lines 26-32. The method finally recites determining and storing at least one source-specific transformation action for the determined transformation errors in a memory **20** (FIG. 4), and identifying and transforming transformation errors in other transformed source-identified texts based on the source-specific transformation actions in the memory. Exemplary transformation actions such as repair are described on page 24, lines 8-17.

A processor **15**, as shown in FIG. 4 of the specification, performs the steps of determining a source-identified training corpus, determining a verified first source-mode transformation, determining an un-verified second source-mode transformation, determining at least one transformation error, determining and storing at least one source-specific transformation action, and identifying and transforming transformation errors in other transformed source-identified texts based on the source-specific transformation actions in the memory.

The invention of <u>claim 81</u> is directed to computer readable storage medium comprising computer readable program code embodied on the computer readable storage medium. The computer readable program code is processable to program a

computer **15** to determine at least one predictive model for a linked event detection system **100** by executing the following recited steps. Claim 81 recites steps, as shown in FIG. 2, for determining source-identified training stories, determining inter-story similarity vectors in a memory **20** (FIG. 4) for at least one story-pair, determining link label information for the at least one story-pair, and determining and storing at least one predictive model in the memory based on the inter-story similarity vectors and the link label information. The link label information indicates the existence of at least one link between a pair of stories in the source-identified training stories and that the linked source-identified stories are related to the same event.

A processor **15**, as shown in FIG. 4 of the specification, performs the step of determining source-identified training stories and the step of determining link label information as described on page 20, line 20 – page 21, line 3. The step for determining inter-story similarity vectors is performed by an inter-story similarity determining circuit **40** as described on page 21, line 28 – page 22, line 30. The step for determining and storing at least one predictive model is performed by a predictive model determining circuit **50** as described on page 20, lines 20-33.

The invention of claim 82 is directed to computer readable storage medium comprising computer readable program code embodied on the computer readable storage medium. The computer readable program code is processable to program a computer **15** to determine at least one predictive model for a linked event detection system **100**. Each of the instructions limitations in claim 82 recites a §112, 6th paragraph, means-plus-function limitation and the disclosed structures, materials, or acts described in the specification that correspond to the claimed step are identified with reference to FIG. 4. The computer readable program code includes instructions for

determining source-identified training stories (processor **15**, page 20, line 20 – page 21, line 3), instructions for determining inter-story similarity vectors in a memory **20** for at least one story-pair (circuit **40**, page 21, line 28 – page 22, line 30), instructions for determining link label information (processor **15**, page 20, line 20 – page 21, line 3) for the at least one story-pair, and instructions for determining and storing at least one predictive model in the memory based on the inter-story similarity vectors and the link label information (circuit **50**, page 20, lines 20-33). The link label information indicates the existence of at least one link between a pair of stories in the source-identified training stories and that the linked source-identified stories are related to the same event.

The invention of claim 83 is directed to computer readable storage medium comprising computer readable program code embodied on the computer readable storage medium. The computer readable program code is processable to program a computer **15** to detect linked events by executing program steps. Claim 83 recites steps, as shown in FIG. 3, for determining source-identified training stories, determining inter-story similarity vectors in a memory **20** (FIG. 4) for at least one story-pair of the source-identified training stories, determining at least one predictive model in the memory for link detection, determining a link between the story-pairs based on the predictive model and the inter-story similarity vectors, and displaying the link on a computer **300** or storing the link in an information repository **200**. The link indicates that the story-pairs are related to the same event.

A processor **15**, as shown in FIG. 4 of the specification, performs the step of determining source-identified training stories as described on page 20, line 20 – page 21, line 3. The step for determining inter-story similarity vectors is performed by an inter-

story similarity determining circuit **40** as described on page 21, line 28 – page 22, line 30. The step for determining at least one predictive model is performed by a predictive model determining circuit **50** as described on page 20, lines 20-33. The step for determining a link between the story-pairs is performed by a link determining circuit **55** as described on page 22, lines 20-30. The step for displaying the link on a computer or storing the link in an information repository is performed by the link determining circuit.

The invention of claim 84 is directed to directed to computer readable storage medium comprising computer readable program code embodied on the computer readable storage medium. The computer readable program code is processable to program a computer **15** to detect linked events. Each of the instructions limitations in claim 84 recites a §112, 6th paragraph, means-plus-function limitation and the disclosed structures, materials, or acts described in the specification that correspond to the claimed step are identified with reference to FIG. 4. The computer readable program code includes instructions for determining source-identified training stories (processor **15**, page 20, line 20 – page 21, line 3), instructions for determining inter-story similarity vectors in a memory **20** for at least one story-pair of the source-identified training stories (circuit **40**, page 21, line 28 – page 22, line 30), instructions for determining at least one predictive model in the memory for link detection (circuit **50**, page 20, lines 20-33), instructions for determining a link between the story-pairs based on the predictive model and the inter-story similarity vectors (link determining circuit **55**, page 22, lines 20-30), and instructions for displaying the link on a computer **300** or storing the link in an information repository **200** (performed by the link determining circuit **55**). The link indicates that the story-pairs are related to the same event.

V. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

The following grounds of rejection are presented for review:

1) Whether claims 1-2, 20-21, 39-40, 58-59, and 81-84 are unpatentable under 35 U.S.C. §103(a) over U.S. Patent No. 6,606,620 issued to Sundaresan et al. (hereinafter Sundaresan) in view of U.S. Patent No. 5,835,905 issued to Pirolli et al. (hereinafter Pirolli) and further in view of U.S. Patent No. 6,961,954 issued to Maybury et al. (hereinafter Maybury).

2) Whether claims 77-80 are unpatentable under 35 U.S.C. §103(a) over Wayne, Charles, "Topic Detection and Tracking in English and Chinese" (hereinafter Wayne) in view of U.S. Patent Application No. 2005/0289463, filed June 23, 2004 by Wu et al. (hereinafter Wu), and in particular, whether the Wu reference is a proper reference.

3) Whether remaining dependent claims 3-19, 22-38, 41-57, 60-76 and 85-88 are distinguished.

VI. ARGUMENT

Applicants re-state herein the arguments set forth in the Appeal Brief submitted December 29, 2008, prior to the most recent Office Action of April 10, 2009, along with new arguments directed to newly set forth rejections of claims 77-88 under 35 U.S.C. §§ 101 and 103.

A. Claims 1-2, 20-21, 39-40, 58-59, and 81-84 (and Remaining Dependent Claims) Would Not Have Been Obvious Over Sundaresan in View of Pirolli and Further in View of Maybury

1. Claim 1

a. Neither Sundaresan Nor Combination of References Teach Use of Source Identification

In the Amendment After Final Rejection filed on January 21, 2008 Applicants argued that the Office Action failed to show where the Sundaresan reference makes any reference to the use of source information for the training documents. It was also argued that the Sundaresan reference appears to be silent on the subject of source information for the training documents, but instead describes only word or term-based vectors. In the non-final Office Action mailed July 28, 2008, in the Response to Arguments, it is asserted that Sundaresan teaches in col. 6, lines 48-49 that web pages are identified by URL and come from web sites (col. 6, lines 65-68) that are associated with particular domain names and include the content of a particular organization. However, neither the Office Action does not show where Sundaresan describes using the URL or even the domain name in embodiments taught by Sundaresan. To the contrary, the URL and domain name described in col. 6, lines 48-49 and lines 65-67 are only being discussed in a definition of background explanation section (see col. 5, lines 44-47) of the Sundaresan reference and are not further discusses in the reference

except to show that the result if a user search on the Internet may include a list of URLs (col. 7, lines 40-42). The Examiner has not shown where the Sundaresan reference, or the Sundaresan reference in combination with the Pirolli and Maybury references, describes utilization of the source information for a document. Merely knowing that a document had a source, or knowing that a document is being retrieved from a source, does not teach or suggest utilization of the source identification as part of determining predictive models for linked event detection. To the contrary, however, the source information of a document is clearly a recited feature in the limitations of claim 1 of the present application, i.e., source-identified.

b. Combination of Sundaresan and Pirolli Does Not Teach Determining Inter-Story Similarity Vectors for Source-Identified Training Stories

The Office Action mailed July 28, 2008 admits that Sundaresan does not teach determining inter-story similarity vectors for at least one story-pair, however, the Office Action asserts that Pirolli teaches determining inter-story similarity vectors, with reference to col. 7, lines 53-65. However, the vectors taught by Pirolli are clearly word-based and <u>do not include source identification information</u>. As described on page 24, line 29 – page 25, line 11 of the present application, the source characteristics used for source identification are associated with a source which may be a CNN, ABC, NBC, Aljazeera or CTV television broadcast, the text of a Reuters newswire service story, an article in the Wall Street Journal or any other known or later developed information source.

The Office Action, however, asserts that because Sundaresan teaches source, that the broadest interpretation of the limitation is covered by combining the references.

Applicants submit that the Examiner is reading features into the references that are neither taught nor suggested as described above under "Sundaresan Does Not Teach Use of Source Identification." The limitations of claim 1 clearly recite determining inter-story similarity vectors <u>for source-identified</u> training stories.

c. Neither Sundaresan Nor Combination of References Teach Determining Link label Information for Source-Identified Training Stories

Applicants submit that Sundaresan does not teach determining link label information for Source-Identified Training Stories. The Office Action, in the Response to Arguments, asserts that classifying documents into categories means that the documents in a particular directory are linked by similarity of terms or concept. However, classifying documents into directories based on similarity of terms or concept is a much broader categorization than determining that two stories are linked to the same event. The Examiner has not shown where the Sundaresan reference, or the Sundaresan reference in combination with the Pirolli and Maybury references, teaches or describes determining that stories are related to the same event. Claim 1, however, clearly recites a limitation for determining link-label information that indicates that the source-identified stories are related to the same event.

d. Neither Maybury Nor Combination of References Teach Indicating Existence of Stories Linked to the Same Event

Applicants also submit that Maybury does not teach indicating the existence of stories linked to the same event. The Office Action, in the Response to Arguments, asserts that Maybury teaches a system that finds interrelated stories using segmentation (col. 19, lines 33-38). Contrary to the Examiner's interpretation of

Maybury, i.e., finding interrelated stories using segmentation, the Maybury reference teaches segmenting a given story into segments for, e.g., more timely and efficient communication and storage of multimedia data (col. 2, lines 41-67). In fact, col. 19, lines 33-38 cited by the Examiner teach exactly the opposite of finding stories linked to the same event. The cited text describes how the "system over-generated story segments, which increased the number of stories that were found. In three cases of over segmentation, a story crossed a commercial boundary and was broken into two individual stories. In one case, a story consisted of two sub-stories, the Peruvian Army and the Peruvian Economy." As evidenced by this cited section of the Maybury reference, it seems the goal is to separate an individual document into multiple documents based on the subject matter, rather than finding separate stories linked to the same event. As described above, claim 1 clearly recites a limitation for determining link label information indicating a link between a pair of stories, and that the linked source-identified stories are related to the same event.

For at least the reasons set forth above, Applicants submit that neither the Maybury reference, nor the Maybury reference in combination with the Pirolli and Sundaresan references, teach the recited limitation for determining link-label information that indicates the source-identified stories are related to the same event.

2. Claims 20, 39, 58 and 81-84

Each of claims 20, 39, 58 and 81-84 recites limitations similar to the above-discussed limitations of claim 1. Therefore, each of the arguments discussed above with reference to claim 1 apply as well to claims 20, 39, 58 and 81-84. Consequently,

Applicants direct the Board's attention to the arguments for claim 1 and have not repeated the arguments here.

3. Claim 2

a. Neither Sundaresan Nor Combination of References Teach Use of Inter-Story Similarity Metrics and Source-Pair Statistics for Determining Inter-Story Similarity Vectors

Applicants note that the Office Action mailed July 28, 2008 asserts that Sundaresan, as modified, is directed to a step of determining inter-story similarity vectors comprising the steps of determining at least one inter-story similarity metric for the story-pairs (Sundaresan, col. 4, lines 9-25), and determining at least one source-pair statistic for the at least one story-pair (Sundaresan, col. 10, lines 15-17). However, the gathered statistical information described by Sundaresan in col. 10, lines 15-17 is term-based and does not include source identification data. For example, col. 10, lines 9-12, Sundaresan describes calculating the frequency and distribution of terms. This is further made clear in lines 17-22 of col. 10: "The statistics are calculated by combining all the documents of a given type together in a meaningful fashion. In particular, the modeling sub-module 415 combines the individual vectors in the class by adding them together and normalizing the result. Term frequencies may be normalized at any level from the uppermost (document level) to the lowest sub-vector." The described step of adding individual vectors is merely adding previously determined term-based vectors and normalizing the results. Contrariwise, claim 2 of the present application recites a limitation wherein inter-story similarity vectors include source-pair statistics which include source identification data. For example, FIG. 6 of the present application shows the steps of determining source-pair similarity statistics which comprise a component of

the inter-story similarity vectors. The process is further described on page 24, line 21 - page 25, line 26 of the present application. It should be noted that in step S1030, the source pair statistics are determined based on the source characteristics of the stories in the source-pair. Specific source pair statistics are maintained for each identified source pair.

Unlike the gathered statistical information of Sundaresan, the recited source pair statistics are not term based. As described on page 24, line 29 – page 25, line 11, the source characteristics upon which the recited source pair statistics are based are associated with a source which may be a CNN, ABC, NBC, Aljazeera or CTV television broadcast, the text of a Reuters newswire service story, an article in the Wall Street Journal or any other known or later developed information source. The source characteristics associated with each source in a source-pair are used to select source-pair similarity statistics from the source hierarchy. The source hierarchy may be based on source characteristics such as source language, input mode and the like. An English radio broadcast captured using automatic speech recognition may be associated with an "English" language source characteristic and an "ASR" input mode source characteristic. A Chinese text translated into English may be associated with a "Chinese" source language characteristic and a "TEXT" input mode characteristic. The two stories thus form a story pair having "English:ASR" and "Chinese:TEXT" source pair characteristics.

Use of the above-described source-pair statistics as a component of similarity vectors, in combination with inter-story similarity vectors, as recited in claim 2, are neither taught nor suggested in the cited references which describe only word/term-

based vectors. Thus, the cited references, either individually or combined, do not teach the inter-story similarity vectors recited in claim 2.

For at least the reasons set forth above, Applicants submit that neither the Sundaresan reference, nor the Sundaresan reference in combination with the Pirolli and Maybury references, teach the recited limitation for determining link-label information that indicates the source-identified stories are related to the same event.

4. Claims 21, 40, and 59

Each of claims 21, 40, and 59 recites limitations similar to the above-discussed limitations of claim 2. Therefore, each of the arguments discussed above with reference to claim 2 apply as well to claims 21, 40, and 59. Consequently, Applicants direct the Board's attention to the arguments for claim 2 and have not repeated the arguments here.

5. Claims 3-19, 22-38, 41-57, 60-76, and 85-88

Each of claims 3-19, 22-38, 41-57, 60-76, and 85-88 depend from and further define the corresponding distinguished independent claims.

B. Claims 77-80 Recite Patentable Subject Matter

The Office Action mailed April 10, 2009 asserts claims 77-80 are directed to unpatentable subject matter.

35 U.S.C. §101 recites four categories of patent-eligible subject matter: processes, machines, manufactures, and compositions of matter. Independent claim 77 is clearly directed to a process. A process claim may be rejected under §101 if it claims "laws of nature, natural phenomena, [or] abstract ideas." *In re Bilski*, 545 F.3d 943 (Fed. Cir. 2008) (en banc). The court in Bilski affirmed that "a claimed process is surely

patent-eligible under §101 if: (1) it is tied to a particular machine or apparatus, or (2) it transforms a particular article into a different state or thing."

Independent claim 77 recites a method of determining a stopword list that includes identifying and transforming transformation errors in transformed source-identified texts based on source-specific transformation actions in memory. The process of transforming of transformation errors in transformed source-identified texts transforms particular articles (source-identified texts), and thus is statutory under §101. As acknowledged by the *Interim Examination Instructions for Evaluating Subject Matter Eligibility Under 35 U.S.C. §101* of August 24, 2009:

> An article can be electronic data that represents a physical object or substance. ... The data should be more than an abstract value. Data can be specifically identified by indicating what the data represents, the particular type or nature of the data, and/or how or from where the data was obtained. (page 5, next-to-last paragraph).

The source-identified texts recited in independent claim 77 are considered articles because they are electronic data that represents a physical text, such as an news article or story. The source-identified texts are transformed through the process of transforming transformation errors in the texts so that the transformation errors (which may be caused by a bad language translation) are corrected. After the transformation errors in the texts are transformed, the texts contain corrected translation errors. Thus, the articles (source-identified texts) are transformed by the action of transforming transformation errors in transformed source-identified texts based on the source-specific transformation action in memory. As acknowledged by the *Interim Examination Instructions*:

> for data, mathematical manipulation *per se* has not been deemed a transformation; but, transformation of electronic data has been found

> when the nature of the data has been changed such that it has a different function or is suitable for a different use. (page 5, last paragraph to page 6, paragraph 1).

The source-identified texts of independent claim 77 are changed such that they are suitable for a different use. After the transformation of the transformation errors in the source-identified texts, the source-identified texts can then be displayed, printed, or otherwise processed in an environment where transformation errors are undesirable. Additionally, the transformation of transformation errors in source-identified texts in independent claim 77 imposes a meaningful limit on the claim's scope, and thus renders the claim patentable under 35 U.S.C. §101.

Accordingly, it is submitted that independent claim 77, and dependent claims 78-80 recite patentable subject matter and reconsideration thereof is respectfully requested.

C. Claims 77-80 Would Not Have Been Obvious Over Wayne in View of Brown

1. Claim 77

a. Neither Wayne Combined With Brown Nor Combination of References Teach Transforming Transformation Errors in Source-Identified Texts

The Office Action mailed April 10, 2009 asserts that claims 77-80 are unpatentable under 35 U.S.C. §103(a) over Wayne in view of Brown. The Office Action cites Brown as teaching transforming transformation errors in source-identified texts. (Office Action, page 32, paragraph 4). However, Applicants would like to note that the cited portion of Brown explicitly teaches away from this transforming source-identified texts:

> Each test document then has the weights of the stopwords for the cluster containing it set to zero in its term vector prior to computing the cosine similarity, effectively removing the stopwords without actually altering the document stored in memory. (Brown, page 2, column 1, lines 2-6; emphasis added).

Thus, the cited portion of Brown never actually transforms transformation errors in transformed source-identified texts based on source-specific transformation actions in memory. Additionally, Wayne does not transform transformation errors in source-identified texts either. Therefore, the combination of Wayne and Brown does not address each and every limitation of independent claim 77. Accordingly, for at least this reason, it is submitted that independent claim 77 and claims dependent thereon (78-80) distinguish patentably and unobviously over the combination of Wayne and Brown.

CONCLUSION

For all of the reasons discussed above, it is respectfully submitted that the rejections are in error and that independent claims 1, 20, 39, 58, 77, and 81-84 are in condition for allowance. Applicants submit also that each of the remaining dependent claims 2-19, 21-38, 40-57, 59-76, 78-80, and 85-88, by reason of dependence from their base independent claims, are also in condition for allowance. For all of the above reasons, Appellants respectfully request this Honorable Board to reverse the rejections of claims 1-88.

Respectfully submitted,

Mark S. Svat
Registration No. 34,261

FAY SHARPE LLP
1100 Superior Avenue – Seventh Floor
Cleveland, Ohio 44114-2579
Telephone: (216) 861-5582

Dated: September 10, 2009

APPENDICES

VII. CLAIMS APPENDIX:

Claims involved in the Appeal are as follows:

1. A computer-implemented method of determining predictive models for a linked event detection system comprising the steps of:

determining source-identified training stories;

determining inter-story similarity vectors in a memory for at least one story-pair of the source-identified training stories;

determining link label information for the at least one story-pair, the link label information indicating the existence of at least one link between a pair of stories in the source-identified training stories and that the linked source-identified stories are related to the same event; and

determining and storing at least one predictive model in the memory based on the inter-story similarity vectors and the link label information.

2. The method of claim 1, wherein the step of determining inter-story similarity vectors comprises the steps of:

determining at least one inter-story similarity metric for the story-pairs; and

determining at least one source-pair statistics for the at least one story-pair.

3. The method of claim 2, wherein determining inter-story similarity vectors further comprise the step of normalizing the inter-story similarity metric based on the source-pair statistics.

4. The method of claim 2, wherein determining inter-story similarity vectors further comprise the step of incrementally normalizing the inter-story similarity metric based on the source-pair statistics.

5. The method of claim 2, wherein the inter-story similarity metric is normalized based on at least one of subtraction and division.

6. The method of claim 2, wherein the inter-story similarity metric is at least one of a probability based similarity metric and a Euclidean based similarity metric.

7. The method of claim 6, wherein the probability based inter-story similarity metric is at least one of a Hellinger, a Tanimoto and a clarity distance based metric.

8. The method of claim 6, wherein the Euclidean based inter-story similarity metric is a cosine-distance based metric.

9. The method of claim 1, further comprising the step of transforming the source-identified training stories.

10. The method of claim 9, wherein transforming the source-identified training stories is at least one of translating, transcribing and linguistically transforming.

11. The method of claim 2, wherein the inter-story similarity metrics are based on terms in at least one source-identified term frequency-inverse story frequency models.

12. The method of claim 11, wherein the terms in source-identified term frequency-inverse story frequency models are based on language.

13. The method of claim 11, wherein determining terms comprises the steps:

determining a reference language; and

determining reference language and non-reference language terms.

14. The method of claim 2, wherein the at least one inter-story similarity metric is normalized based on at least one of a source-pair identified similarity statistic.

15. The method of claim 1, wherein the at least one predictive model is at least one of: a classifier, a support vector machine, a decision tree and a Naive-Bayes classifier.

16. The method of claim 2, wherein at least one of the source-pair similarity statistics are determined based on a source hierarchy.

17. The method of claim 16 wherein the source hierarchy is determined based on at least one source characteristic.

18. The method of claim 16 wherein the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre characteristic, a source name characteristic and a transformation characteristic.

19. The method of claim 16 wherein the source-pair similarity statistic for a new source is determined based on at least one source characteristic of the new source.

20. A linked event detection training system comprising:

an input/output circuit;

a memory;

a processor that receives source-identified training stories and associated link label information for at least one story-pair via the input/output circuit, the link label information indicating the existence of at least one link between a pair of stories in the source-identified training stories and that the linked source-identified stories are related to the same event;

an inter-story similarity vector determining circuit that determines inter-story similarity vectors in the memory for at least one story-pair of the source-identified training stories; and

a predictive model determining circuit that determines and stores at least one predictive model in the memory based on the inter-story similarity vectors and the link label information.

21. The system of claim 20, wherein the inter-story similarity vector determining circuit is comprised of:

a similarity metric determining circuit that determines at least one inter-story similarity metric for the at least one story-pair; and

a similarity statistics determining circuit that determines at least one source-pair statistic for the at least one story-pair.

22. The system of claim 21, wherein the inter-story similarity vector determining circuit normalizes the inter-story similarity metric based on the source-pair statistics.

23. The system of claim 21, wherein the inter-story similarity vector determining circuit incrementally normalizes the inter-story similarity metric based on the source-pair statistics.

24. The system of claim 21, wherein at least one of the inter-story similarity metrics is normalized based on at least one of a subtraction and a division operation.

25. The system of claim 21, wherein at least one of the inter-story similarity metrics is at least one of a probability based similarity metric and a Euclidean based similarity metric.

26. The system of claim 25, wherein the probability based inter-story similarity metric is at least one of a Hellinger, a Tanimoto and a clarity distance based metric.

27. The system of claim 25, wherein the Euclidean based inter-story similarity metric is a cosine-distance based metric.

28. The system of claim 20, wherein the source-identified training stories are transformed.

29. The system of claim 28, wherein transforming the source-identified training stories is at least one of translating, transcribing and linguistically transforming.

30. The system of claim 20, wherein the inter-story similarity metrics are based on terms in at least one source-identified term frequency-inverse story frequency model.

31. The system of claim 30, wherein the terms in the source-identified term frequency-inverse story frequency models are based on language.

32. The system of claim 30, wherein the processor determines terms based on a reference language; and determining reference language and non-reference language terms.

33. The system of claim 21 wherein the at least one inter-story similarity metric is normalized based on at least one of a source-pair identified similarity statistic.

34. The system of claim 20, wherein the at least one predictive model is at least one of: a classifier, a support vector machine, a decision tree and a Naive-Bayes classifier.

35. The system of claim 21, wherein the source-pair identified similarity statistic is determined based on a source hierarchy.

36. The system of claim 35, wherein the source hierarchy is determined based on at least one of a source characteristic.

37. The system of claim 35, wherein the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre characteristic, a source name characteristic and a transformation characteristic.

38. The system of claim 35, wherein the source-pair similarity statistic for a new source is determined based on at least one source characteristics of the new source.

39. A computer-implemented method of linked event detection comprising the steps of:

determining source-identified stories;

determining inter-story similarity vectors in a memory for the story-pairs of the source-identified stories;

determining at least one predictive model in the memory for link detection;

determining a link between the story-pairs based on the predictive model and the inter-story similarity vector; and

displaying the link on a computer or storing the link in an information repository, the link indicating the story-pairs are related to the same event.

40. The method of claim 39, wherein the step of determining inter-story similarity vectors comprises the steps of:

determining at least one inter-story similarity metric for each story-pair; and

determining source-pair statistics for the story-pairs.

41. The method of claim 40, wherein determining inter-story similarity vectors further comprise the step of normalizing the inter-story similarity metric based on the source-pair statistics.

42. The method of claim 40, wherein determining inter-story similarity vectors further comprise the step of incrementally normalizing the inter-story similarity metric based on the source-pair statistics.

43. The method of claim 40, wherein the inter-story similarity metric is normalized based on at least one of subtraction and division.

44. The method of claim 40, wherein the inter-story similarity metric is at least one of a probability based similarity metric and a Euclidean based similarity metric.

45. The method of claim 44, wherein the probability based inter-story similarity metric is at least one of a Hellinger, a Tanimoto and a clarity distance based metric.

46. The method of claim 44, wherein the Euclidean based similarity metric is a cosine-distance based metric.

47. The method of claim 39, further comprising the step of transforming the source-identified training stories.

48. The method of claim 47, wherein transforming the source-identified training stories is at least one of translating, transcribing and linguistically transforming.

49. The method of claim 40, wherein the inter-story similarity metrics are based on terms in at least one source-identified term frequency-inverse story frequency models.

50. The method of claim 49, wherein the terms in source-identified term frequency-inverse story frequency models are based on language.

51. The method of claim 49, wherein determining terms comprises the steps:

determining a reference language; and

determining reference language and non-reference language terms.

52. The method of claim 40, wherein the at least one inter-story similarity metric is normalized based on at least one of a source-pair identified similarity statistic.

53. The method of claim 39, wherein the at least one predictive model is at least one of: a classifier, a support vector machine and a decision tree, a Naive-Bayes-classifier.

54. The method of claim 40, wherein the source-pair identified similarity statistic is determined based on a source hierarchy.

55. The method of claim 54, wherein the source hierarchy is determined based on at least one of a source characteristic.

56. The method of claim 54, wherein the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre characteristic, a source name characteristic and a transformation characteristic.

57. The method of claim 54, wherein the source-pair similarity statistic for a new source is determined based on at least one source characteristics of the new source.

58. A linked event detection system comprising:

an input/output circuit;

a memory;

a processor that receives source-identified stories via the input/output circuit;

an inter-story similarity vector determining circuit that determines inter-story similarity vectors in the memory for the story-pairs of the source-identified stories; and

a link determining circuit that determines and displays on a computer or stores in an information repository, links between story-pairs based on a predictive model in the memory and the inter-story similarity vectors, the links indicating the story-pairs are related to the same event.

59. The system of claim 58, wherein the inter-story similarity vector determining circuit is comprised of:

a similarity metric determining circuit that determines at least one inter-story similarity metric for the story-pairs; and

a similarity statistics determining circuit that determines source-pair statistics for the story-pairs.

60. The system of claim 59, wherein the inter-story similarity vector determining circuit normalizes the inter-story similarity metric based on the source-pair statistics.

61. The system of claim 59, wherein the inter-story similarity vector determining circuit incrementally normalizes the inter-story similarity metric based on the source-pair statistics.

62. The system of claim 59, wherein at least one of the inter-story similarity metrics is normalized based on at least one of a subtraction and a division operation.

63. The system of claim 59, wherein at least one of the inter-story similarity metrics is at least one of a probability based similarity metric and a Euclidean based similarity metric.

64. The system of claim 63, wherein the probability based inter-story similarity metric is at least one of a Hellinger, a Tanimoto and a clarity distance based metric.

65. The system of claim 63, wherein the Euclidean based inter-story similarity metric is a cosine-distance based metric.

66. The system of claim 58, wherein the source-identified training stories are transformed.

67. The system of claim 66, wherein transforming the source-identified training stories is at least one of translating, transcribing and linguistically transforming.

68. The system of claim 59, wherein the inter-story similarity metrics are based on terms in at least one source-identified term frequency-inverse story frequency model.

69. The system of claim 68, wherein the terms in the source-identified term frequency-inverse story frequency models are based on language.

70. The system of claim 68, wherein the processor determines terms based on a reference language; and non-reference language terms.

71. The system of claim 59, wherein the at least one inter-story similarity metric is normalized based on at least one of a source-pair identified similarity statistic.

72. The system of claim 58, wherein the predictive model is at least one of: a classifier, a support vector machine and a decision tree, a Naive-Bayes classifier.

73. The system of claim 59, wherein the source-pair identified similarity statistic is determined based on a source hierarchy.

74. The system of claim 73, wherein the source hierarchy is determined based on at least one of a source characteristic.

75. The system of claim 73, wherein the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre characteristic, a source name characteristic and a transformation characteristic.

76. The system of claim 73, wherein the source-pair similarity statistic for a new source is determined based on at least one source characteristics of the new source.

77. A method of determining a stopword list comprising the steps of:

determining a source-identified training corpus of text information;

determining a verified first source-mode transformation of the source-identified training corpus text from a first mode to a second mode based on at least one of a verified transcription and a verified translation;

determining an un-verified second source-mode transformation of the source-identified training corpus text from a first mode to a second mode;

determining at least one transformation error associated with distribution differences between the first and second transformations and identified sources;

determining and storing at least one source-specific transformation action for the determined transformation errors in a memory; and

identifying and transforming transformation errors in other transformed source-identified texts based on the source-specific transformation actions in the memory.

78. The method of claim 77, wherein the first mode is at least one of a text source, an optical character recognition source and an automatic speech recognition source.

79. The method of claim 77, wherein the second mode is at least one of a text source, an optical character recognition source and an automatic speech recognition source.

80. The method of claim 77, wherein the source-specific transformation is at least one of a removal, a repair and a normalization transformation.

81. Computer readable storage medium comprising: computer readable program code embodied on the computer readable storage medium, the computer readable program code processable to program a computer to determine at least one predictive model for a linked event detection system by executing steps comprising:

determining source-identified training stories;

determining inter-story similarity vectors in a memory for at least one story-pair;

determining link label information for the at least one story-pair of the source-identified training stories, the link label information indicating training stories related to the same event; and

determining and storing at least one predictive model in the memory based on the inter-story similarity vectors and the link label information.

82. Computer readable storage medium comprising: computer readable program code embodied on the computer readable storage medium, the computer readable program code processable to program a computer to determine at least one predictive model for a linked event detection system, the computer readable program code comprising:

instructions to determine source-identified training stories;

instructions to determine inter-story similarity vectors in a memory for at least one story-pair of the source-identified training stories;

instructions to determine link label information for the at least one story-pair, the link label information indicating training stories related to the same event; and

instructions to determine and store at least one predictive model in the memory based on the inter-story similarity vectors and the link label information.

83. Computer readable storage medium comprising: computer readable program code embodied on the computer readable storage medium, the computer readable program code processable to program a computer to detect linked events by executing steps comprising :

determining source-identified stories;

determining inter-story similarity vectors in a memory for the at least one story-pair of the source-identified stories;

determining at least one predictive model in the memory for link detection; and

determining a link between story-pairs based on the at least one predictive model and the inter-story similarity vectors, the link indicating the story-pairs are related to the same event; and

displaying the link on a computer or storing the link in an information repository.

84. Computer readable storage medium comprising: computer readable program code embodied on the computer readable storage medium, the computer readable program code processable to program a computer to detect linked events, the computer readable program code comprising:

instructions to determine source-identified stories;

instructions to determine inter-story similarity vectors in a memory for the at least one story-pair of the source-identified stories;

instructions to determine at least one predictive model in the memory for link detection;

instructions to determine a link between story-pairs based on the predictive model and the inter-story similarity vectors, the link indicating the story-pairs are related to the same event; and

instructions to display the link on a computer or store the link in an information repository.

85. The method of claim 2, wherein determining at least one source-pair statistic for the at least one story-pair is based on at least one of a similarity metric and a statistic associated with the metric.

86. The system of claim 21, wherein determining at least one source-pair statistic for the at least one story-pair is based on at least one of a similarity metric and a statistic associated with the metric.

87. The method of claim 39, wherein at least one of the predictive models is a trained predictive model.

88. The system of claim 58, wherein at least one of the predictive models is a trained predictive model.

VIII. EVIDENCE APPENDIX

A copy of each of the following items of evidence relied on by the Appellant [and/or the Examiner] is attached:

NONE